



Performance assessment of probe-based Raman spectroscopy systems for biomedical analysis

SEAN FITZGERALD,^{1,2}  ERIC MARPLE,³ AND ANITA MAHADEVAN-JANSEN^{1,2}

¹Vanderbilt Biophotonics Center, Nashville, TN 37232, USA

²Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37232, USA

³EmVision LCC, 1471 F Road, Loxahatchee, FL 33470, USA

Abstract: We present a methodology for evaluating the performance of probe-based Raman spectroscopy systems for biomedical analysis. This procedure uses a biological standard sample and data analysis approach to circumvent many of the issues related to accurately measuring and comparing the signal quality of Raman spectra between systems. Dairy milk is selected as the biological standard due to its similarity to tissue spectral properties and because its homogeneity eliminates the dependence of probe orientation on the measured spectrum. A spectral dataset is first collected from milk for each system configuration, followed by a model-based correction step to remove photobleaching artifacts and accurately calculate SNR. Results demonstrate that the proposed strategy, unlike current methods, produces an experimental SNR that agrees with the theoretical value. Four preconfigured imaging spectrographs that share similar manufacturer specifications were compared, showing that their capabilities to detect biological Raman spectra widely differ in terms of throughput and stray light rejection. While the methodology is used to compare spectrographs in this case, it can be adapted for other purposes, such as optimizing the design of a custom-built Raman spectrometer, evaluating inter-probe variability, or examining how altering system subcomponents affects signal quality.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Raman Spectroscopy (RS) is a powerful optical method capable of examining the biochemical content of a sample in a non-destructive and label-free fashion, yielding specific compositional information on a molecular level [1]. Following the advent and broader availability of laser technology, RS has shown great potential to interrogate tissues and biofluids for applications in clinical diagnostics [2]. In many cases, fiber optic probes (FOPs) enhance the versatility of developing health monitoring devices [3] since optical fibers hold distinct advantages in this context (e.g., flexible, small diameter, cheap). Most importantly, they provide the key benefit of decoupling the laser source and signal detection platform from the sampling volume, allowing for remote and in situ RS measurements. In effect, FOPs allow more flexibility in device design, permitting the integration of RS into the instrument channel of standard medical endoscopes [4,5] and even through hypodermic needles [6].

The Raman scattering cross-section of organic molecules is inherently low [7], so dispersive RS systems that are designed for biomedical analysis require careful optical design and high-performance detection components. In practice, enhancing the detection of weak Raman signals can be achieved in three primary ways: (1) high numerical aperture (NA) light collection optics, (2) detectors with high quantum efficiency, and (3) high-throughput spectrographs. Concerning FOP systems, the low NA of fiber optics typically between 0.1–0.3 sets the primary constraint on light-gathering ability. While bundling multiple collection fibers within the FOP can increase the amount of optical signal gathered from the sample, there is an upper limit on the number of fibers that can be efficiently imaged in the RS spectrograph due to off-axis diffraction angles causing a curvature in the spectral image that degrades spectral resolution [8]. Utilizing detectors with

high quantum efficiency and low noise performance can also enhance system sensitivity, and manufacturer-provided specifications help decide between commercially available options. Still, much of the optimization space for improving RS systems rests on optimizing the performance of the spectrograph.

However, assessing the performance of the imaging spectrograph used for biomedical RS is not straightforward because of the specific characteristics of biological Raman spectra that lead to low signal quality. The Raman-scattered component of the detected spectrum is overwhelmed by a large proportion of autofluorescence, typically multiple times stronger than the Raman. Additionally, the weak nature of Raman scattering from biological specimens also makes the spectrograph vulnerable to stray light, which refers to any light reaching the detector that deviates from the intended optical path. Stray light interference may arise from ambient light bleed, backscattering of the excitation laser, or imperfections in optical components within the spectrograph [9]. Because of these competing signal contributions, a spectrograph used for detecting biological Raman spectra requires specific attributes: high-efficiency diffraction gratings, an optical bench with sufficient f-number to capture all the light delivered by the FOP, light-tight construction, and high-performance filters to block the Raman laser line [10]. The manufacturer specifications typically define the spectral range, resolution, and f-number of the spectrograph, but often lack any distinct information about realizable light throughput or stray light rejection. This leaves the researcher or engineer with the challenge of finding the optimal spectrograph design for detecting biological Raman spectra.

When testing a RS system's performance, it is often overlooked that the sample itself affects various properties of the collected optical signal [11]. First, the scattering coefficient of the sample impacts the amount of backscattered laser light and, therefore, stray light entering the spectrometer. Second, the sample's composition influences the proportion of autofluorescence to Raman signal present in the spectrum. Using a biological sample ensures that the testing data accurately captures these aspects of true tissues and will reflect the performance of the system for biomedical analysis. The sample's homogeneity is also an important consideration when comparing multiple RS system configurations, as the repeatability of the optical signal is affected by the placement and orientation of the FOP with respect to the sample. It is crucial to input identical signals across all system configurations to prevent experimental variations from skewing comparison analysis of the tested RS systems.

The signal-to-noise ratio (SNR) is a useful metric to quantify Raman signal quality as it is a measure of the strength of the Raman signal relative to the amount of signal variance due to noise. Yet, the SNR of biological Raman spectra is seldom reported. This is largely because of the photobleaching effect that decreases the autofluorescent background intensity over time, leading to signal instability and challenges in comparing consecutive data frames to estimate signal variance. Reports have attempted to experimentally estimate SNR from a single spectrum by using the variance of a Raman-silent region (e.g., $1800\text{--}1900\text{cm}^{-1}$) [12–15], but this approach requires analyzing the noise over a region of the spectrum that does not reflect the true variance at regions where the Raman peaks lie. This ultimately leads to inaccurate results [16]. For this reason, recent reports analyzing theoretical and experimental Raman SNR utilize non-biological polymers as reference materials that are resistant to photobleaching [17] or avoid experimental calculation of SNR from biological spectra altogether [11].

The challenges in unbiased quantification of signal quality from biological samples point to the motivation of this methodology report. Here, a repeatable and reliable procedure is outlined to assess the performance of RS system configurations by selecting dairy milk as a biological standard sample. By closely emulating the Raman signal of tissues, milk provides a suitable estimate of system data quality for biomedical analysis. Submerging the FOP into this liquid sample also eliminates probe orientation dependence on the data and circumvents experimental variances. The Raman dataset undergoes a model-based correction step to remove

photobleaching artifacts and enables accurate SNR estimation. Results show that this procedure offers improved accuracy over existing methods, where the experimental SNR agrees with the theoretically determined value. This method was then used to compare four preconfigured Raman imaging spectrographs to demonstrate its utility in assessing RS system performance.

2. Materials and methods

2.1. Raman system

Raman datasets were acquired using a narrowband, wavelength-stabilized 785 nm laser diode (II0785MU0350MS, Innovative Photonic Solutions, NJ) and an FOP that was designed for operation at this laser line. The FOP (EmVision LLC, FL) includes a single 300-micron core laser delivery fiber with 0.22 NA surrounded by a ring of seven 300-micron core collection fibers with equivalent NA to couple diffusely reflected light into the spectrometer. The laser fiber is affixed with a bandpass filter centered at the laser line to block extraneous wavelengths from reaching the sample. An annular-shaped 800 nm longpass filter is set atop the proximal end of the collection fibers for first-stage filtering of the back-scattered laser light. The distal end of the collection fibers is linearly arranged at an SMA-terminated fiber coupler that is aligned to the entrance slit of the various spectrographs tested. The optical signal was recorded with a near-infrared optimized and thermoelectrically cooled deep-depleted CCD camera (Blaze 400HR, Teledyne Princeton Instruments, CA), as depicted in Fig. 1(a).

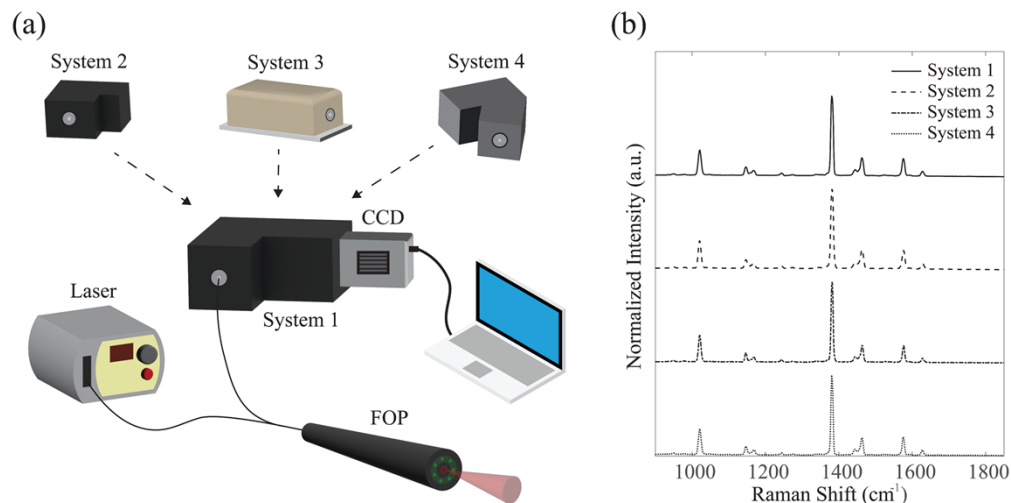


Fig. 1. (a) Schematic of the RS system used to acquire biological spectra, including 785 nm laser diode, charge-coupled device (CCD), fiber optic probe (FOP), and the four preconfigured spectrograph systems. (b) Calibrated and normalized naphthalene spectra, offset for clarity.

The reported methodology was used to compare the performance of four preconfigured Raman spectrographs, which are all designed to resolve Raman spectra at 785 nm excitation with fiber-coupled input and external camera mounting. They share similar optical specifications, including resolution and f-number. An f-number of $f/2.2$ or faster is sufficient for collecting all the light from the 0.22 NA collection fibers within the FOP, which is true for all tested instruments. The spectrographs do, however, differ in the general design of their optical bench and internal components. For example, Spectrographs 1 and 2 utilize a filtering design that decreases the number of internal lenses by butt-coupling the input fibers directly against the entrance slit, thereby minimizing optical losses and the generation of stray light from lens surfaces. These two

instruments, along with Spectrograph 3, use a transmissive grating that is more efficient than the reflective grating used in Spectrograph 4.

2.2. Wavenumber calibration

To map pixel number to relative Raman shift, wavelength dispersion was calibrated with the atomic emission spectrum of a Neon-Argon lamp. Relative wavenumber was then calculated and confirmed with acetaminophen and naphthalene standard materials to ensure that the calibration error was below the system's spectral resolution, which was calculated from the full-width half maximum of an atomic emission line within the Neon-Argon spectrum. As seen from the calibrated and intensity-normalized naphthalene spectra in Fig. 1(b), all spectrographs successfully reproduce Raman data from the chemical standard with comparable resolution.

2.3. Signal-to-noise (SNR) estimation

The SNR of an optical spectrum represents the amount of detectable signal relative to the uncertainty of the true signal value due to experimental and instrumental noise. It is experimentally computed from the Raman peak intensity (S) and the standard deviation (σ) of the peak intensity at a selected wavenumber $\tilde{\nu}$ over multiple acquisitions, defined as the ratio

$$SNR = \frac{S(\tilde{\nu})}{\sigma(\tilde{\nu})} \quad (1)$$

Furthermore, it is possible to determine the theoretical Raman SNR from a combination of independent noise sources and experimental data [11]. The noise is represented by instrumental and experimental sources added in quadrature because they can be regarded as statistically independent [18]. The theoretical SNR equation at wavenumber $\tilde{\nu}$ is given by

$$SNR = \frac{S(\tilde{\nu})}{\sqrt{n_{shot}^2(\tilde{\nu}) + n_{read}^2 + n_{dark}^2}} \quad (2)$$

where S is the Raman peak intensity taken from experimental data, n_{shot}^2 is the variance due to shot noise of detected photons, n_{read}^2 is the variance from detector readout noise, and n_{dark}^2 is the variance of the detector dark current. Noise sources related to laser stability and flicker noise [16] are not considered in this analysis. It is important to consider that photon shot noise is described by Poisson statistics, which states that the magnitude of shot noise is proportional to the square root of the number of detection events [19]. Therefore, shot noise must be estimated from detected photoelectron counts (e^-) as opposed to the analog-to-digital units (ADUs) given by the detector. This conversion is required to estimate shot noise and obtain an accurate theoretical value for Raman SNR.

2.4. Detector characterization

Transforming ADU to absolute photoelectron count is achieved using the camera's amplifier gain. The well established mean-variance method [20] was applied to measure amplifier gain of the detector. A tungsten-halogen lamp was used as a broadband light source, where 500 frames of the smoothly varying spectrum were analyzed to generate the mean-variance plot in Fig. 2(a). The conversion factor between e^- and ADU, or gain, is calculated from the slope of a regression line fitted to this mean-variance relationship for each detector gain setting. The lowest gain setting shows a steep decline in the mean-variance relationship near 40 k ADU (i.e., photoelectron count of 180 ke $^-$ using the gain coefficient of 4.46 e^- /ADU) due to the detector's full well capacity of 180 ke $^-$. So, the Medium gain setting was chosen for further data collection to maintain gain linearity over the full dynamic range.

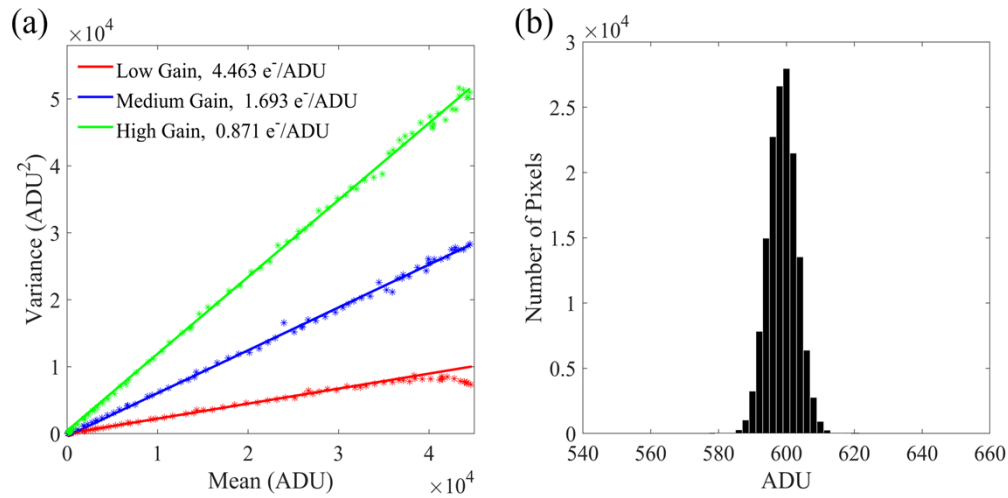


Fig. 2. (a) Gain estimation of the detector using mean-variance plot at three gain settings. (b) Histogram of pixel intensities from a dark measurement to estimate signal-free noise (i.e., read noise and dark current) and DC bias.

To accurately account for all sources of noise contributing to signal variance, it is necessary to include the magnitude of readout noise and dark current in the calculation of the theoretical signal-to-noise ratio (SNR). While readout noise and dark current are provided in the manufacturer specifications of the detector used in this report, these values were validated as follows. The DC bias and signal-free noise were estimated from the mean and standard deviation, respectively, of pixel intensities from a dark measurement recorded using the same detector settings as the experimental data presented in this report (i.e., 2 sec acquisition time, medium gain, -80°C detector cooling). The histogram of dark measurement pixel values is shown in Fig. 2(b), where the bias counts are 600 ADU and the signal-free noise was found to be 4 ADU or 7 e^- . The amplifier gain and signal-free noise intensity both agree with this detector's manufacturer specifications.

2.5. Biological standard

Dairy milk was found to be an advantageous biological standard for the following reasons. It has a similar reduced scattering coefficient to mimic the diffuse reflectance behavior of tissues; a property that has been exploited by using milk as a scattering agent in tissue-mimicking optical phantoms [21,22]. Milk also contains a variety of organic molecules, such as proteins, lipids, and carbohydrates that provide a reasonably similar Raman spectral line shape to that of tissues, as demonstrated in Section 3.1. It is also highly homogenous, which effectively eliminates FOP orientation dependence on the acquired signal. Additionally, milk is relatively inexpensive and easily accessible, making it a convenient choice as a biological standard for researchers who want to assess the performance of probe-based RS systems. Many varieties of milk are commercially available, but this work determined that 2% reduced-fat homogenized milk has a comparable proportion of autofluorescence to Raman scattering intensity that is most similar to tissue spectra.

2.6. Data collection

All spectrographs were tested using the same auxiliary system components, including laser, FOP, and camera, to eliminate their dependence on the measured data. The FOP collection fibers

and camera were aligned onto each instrument and the wavenumber axis was calibrated prior to measurement.

The following camera parameters were set for all measurements: Medium gain ($1.69 \text{ e}^-/\text{ADU}$), 1 Mhz readout speed, vertical hardware binning, -80°C cooling setpoint, and 2 sec acquisition time. Data acquisition was controlled via LightField software (Teledyne Princeton Instruments, CA) for readout from the detector. A laser power of 75 mW from the 785 nm diode was measured at the tip of the FOP before each measurement to ensure consistent irradiance at the sample. A 300 mL solution of dairy milk was added to a beaker and the FOP was dipped 1 cm into the sample and stably positioned by clamping it into a custom chuck. The room lights were turned off during acquisition to minimize any interfering ambient light from entering the FOP. A dataset of 40 consecutive spectra was acquired for a total measurement time of 80 s. The FOP tip was then cleaned with deionized water, and this process is repeated for each spectrograph using a fresh aliquot of milk from the stock.

2.7. Preprocessing

Signal preprocessing was minimally applied to avoid potential influences on the calculation of SNR. The preprocessing steps that were taken will be outlined here, which served to estimate the Raman intensity and remove photobleaching artifacts. Data preprocessing was performed in Matlab R2022b software (Mathworks Inc., Natick, MA, USA).

2.7.1. Baseline estimation

Estimation of the baseline is required to properly define Raman peak intensity and calculate the SNR as it related to the usable Raman signal. This is especially important for biological spectra, where the Raman signal counts are much weaker than non-Raman background counts. The adaptive iteratively reweighted Penalized Least Squares (airPLS) method was applied here to estimate the baseline, which provides good baseline stability in low SNR environments [23].

2.7.2. Photobleaching correction

The effects of photobleaching during continuous laser exposure cause a decrease in autofluorescence that follows a characteristic negative exponential decay [24]. This can be seen within an example milk Raman dataset, where the mean photoelectron count of each spectrum decreases over time (Fig. 3(a)). A subset of these spectra is plotted in Fig. 3(b) to better visualize the direct impact of bleaching, which seems to affect the lower wavenumber region of the spectrum to a greater degree than the higher wavenumber region. Towards the quantification of Raman SNR, this has the deleterious effect of adding false variance within the data. As seen in Fig. 3(c), the standard deviation of photoelectron counts (e^-) is much higher than $\sqrt{e^-}$, which are expected to be proportional for shot-noise limited detection [19]. Therefore, the statistics of the raw spectral dataset contain artificially high signal variance due to photobleaching and would generate a lower experimental SNR when calculated by Eq. (1).

A straightforward and elegant preprocessing algorithm for correcting additive effects in spectral data is Multiplicative Signal Correction (MSC). This model-based correction step is extensively used in near-infrared spectroscopy data to correct variations in sample thickness, atmospheric components, and temperature fluctuations [25]. The MSC algorithm is also useful for Raman data preprocessing for shift correction and removing other measurement-related variations [26–28] but has not yet been reported for photobleaching correction. The MSC model (Guo) is given by

$$I(\tilde{\nu}) = a + b \cdot m(\tilde{\nu}) \quad (3)$$

where the measured spectrum $I(\tilde{\nu})$ is modeled onto a reference spectrum $m(\tilde{\nu})$ that is typically chosen to be the mean of the dataset. The parameters a and b in Eq. (3) are computed through a

first order least-squares fit of each spectrum to the reference. Each photobleaching-corrected spectrum is then obtained according to

$$I_c(\tilde{\nu}) = (I(\tilde{\nu}) - a) / b \quad (4)$$

where $I_c(\tilde{\nu})$ is the corrected Raman spectrum. This procedure is repeated for each measurement in the dataset. The MSC-corrected dataset demonstrates a stable mean (Fig. 3(d)) and effectively shifts all wavenumber regions for their proportional amount of photobleaching decay (Fig. 3(e)). The pixel-wise data statistics in Fig. 3(f) demonstrate that the shot-noise limited relationship between counts and variance is attained, where the standard deviation of counts is now proportional to the square root. This confirms that the MSC procedure does not affect the authentic signal variance caused by instrumental and shot noise sources.

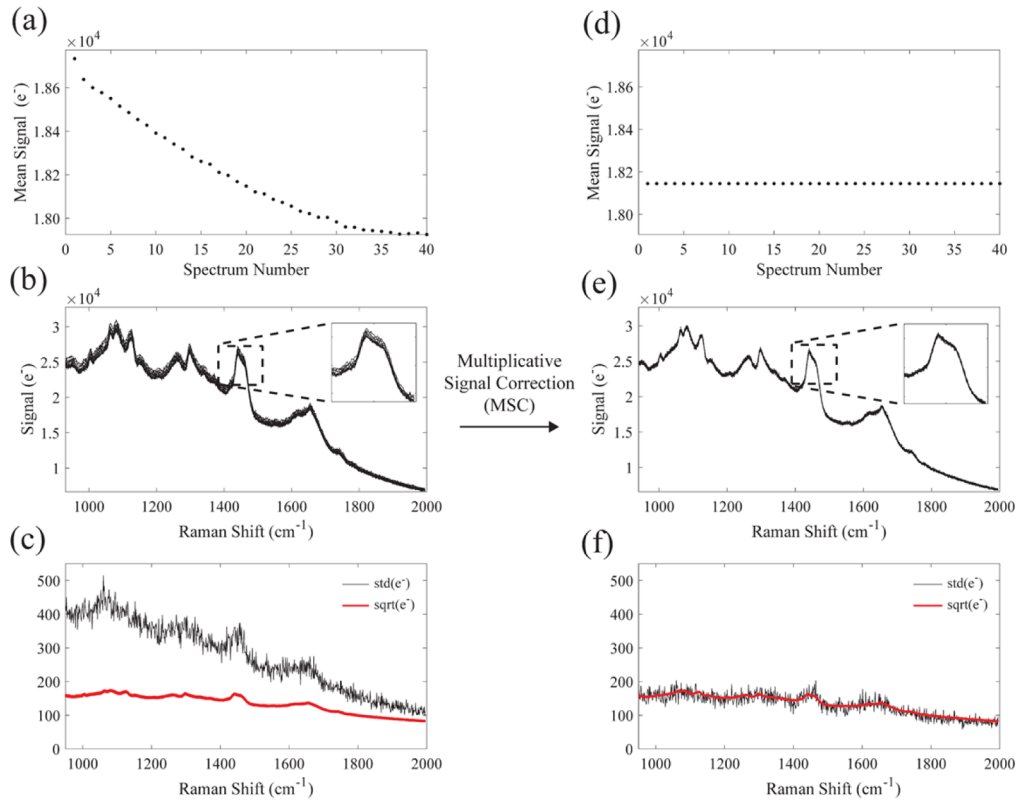


Fig. 3. (a) Mean values of 40 consecutive spectra acquired from dairy milk, demonstrating the effect of photobleaching over consecutive measurements. (b) Plot of a subset of spectra; inset is a zoom of the Raman peak at 1440 cm^{-1} used for SNR calculation. (c) Standard deviation and square root of the dataset on a per-pixel basis. (d-f) Plots the same metrics after MSC correction. All data is presented in photoelectron count (e⁻), calculated using the detector amplifier gain.

3. Results

Before evaluating the performance of various preconfigured Raman spectrographs, the photobleaching correction method using MSC was first validated. The Raman SNR was calculated from experimental data and compared against the theoretical value.

3.1. Experimental SNR validation

Milk Raman spectra were acquired with an FOP following the procedure reported in Section 2.4. The raw data was corrected for the detector DC bias of 600 ADU, as these counts do not relate to detected photoelectrons and should not be included in the estimation of shot noise. The mean spectrum from a dataset of 40 milk spectra and the associated airPLS baseline estimate are plotted in Fig. 4(a), and the background-subtracted mean spectrum is shown in Fig. 4(b). Signal counts are given in both ADU and photoelectron counts, which were converted using the measured detector gain of $1.69 \text{ e}^-/\text{ADU}$. The Raman spectrum of milk is in good agreement with previous reports [29–31].

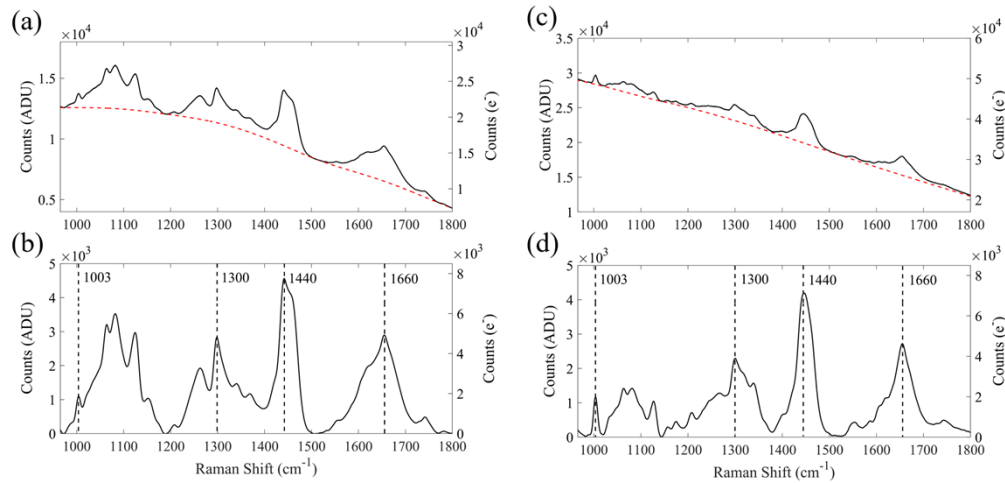


Fig. 4. Raman data from dairy milk and a tissue specimen. (a) Mean spectrum of the milk dataset (black curve) and the airPLS baseline estimate (red dotted curve). (b) The background-subtracted milk spectrum used to estimate Raman intensity, with prominent peaks identified at 1003, 1300, 1440, and 1660 cm^{-1} . Signal intensities are expressed both in analog-to-digital units (ADU) and photoelectron (e^-) counts. (c-d) Plots of the same spectral data from a tissue specimen.

This procedure was repeated with a sample of bovine flesh, and analogous data is plotted in Fig. 4(c)-(d) to demonstrate the similarities in signal characteristics between the biological standard and a true tissue specimen. Milk shares a similar proportion of autofluorescence (Fig. 1a,c) and Raman scattering intensity (Fig. 4(b),d)) as tissue, as well as many characteristic biological spectral features including 1003 cm^{-1} for phenylalanine, 1300 cm^{-1} for CH_2 twisting, 1440 cm^{-1} for CH_2 stretching, and 1660 cm^{-1} for $\text{C}=\text{C}$ stretching [29].

The signal metrics in Table 1 were calculated from the datasets shown in Fig. 4. Total signal intensity and Raman intensity were taken from the 1440 cm^{-1} peak, which is the most prominent Raman band and serves as a benchmark for Raman SNR. To calculate the theoretical Raman SNR, the signal shot noise was calculated as the square root of total signal intensity in photoelectron units. The manufacturer provides the signal-free noise of the detector, including both readout noise and dark current. Still, these instrumental noise sources were empirically confirmed through the readout of a dark frame at equivalent detector settings to those used for Raman measurements (see Section 2.4). Using these experimental and instrumental values, the theoretical Raman SNR values for both milk and the tissue specimen were computed at 1440 cm^{-1} according to Eq. (2).

The experimental Raman SNR values for milk and the tissue specimen were then calculated according to Eq. (1) as the ratio of Raman intensity at 1440 cm^{-1} to the standard deviation of this peak intensity across the 40 spectra within the dataset. The pixel-level standard deviation

Table 1. Calculation of theoretical and experimental SNR for Raman data of biological samples

Signal Metric		Milk	Tissue	Units
Total signal intensity		14900, 25330	24140, 41038	ADU, e-
Raman intensity (1440 cm ⁻¹)		4559, 7750	4333, 7366	ADU, e-
Signal shot noise (1440 cm ⁻¹)		159	203	e-
Signal-free noise sum		7	7	e-
Signal standard deviation	Silent region, 1800-1900cm ⁻¹	45	58	ADU
	Uncorrected, 1440 cm ⁻¹	201	304	ADU
	MSC-corrected, 1440 cm ⁻¹	95	110	ADU
SNR ₁₄₄₀	Theoretical	49	36	a.u.
	Experimental (silent region)	101	75	a.u.
	Experimental (uncorrected)	23	14	a.u.
	Experimental (MSC-corrected)	48	39	a.u.

at 1440 cm⁻¹ was calculated for the uncorrected and MSC-corrected datasets, indicating that photobleaching in the uncorrected dataset overestimates signal variance at the Raman peak. The uncorrected data results in a lower SNR than the theoretical value due to this amplified variance. Signal variance in the Raman-silent region between 1800-1900cm⁻¹ of a single preprocessed spectrum was also calculated to emphasize the issue with this noise estimation method used in previous reports [12–15]. The standard deviation of a silent region underestimates signal variance in both biological samples because of the lower overall signal counts, and consequently lower shot noise, compared to the signal counts at the 1440 cm⁻¹ peak. Using the silent region to estimate signal variance, therefore, generates a higher SNR than the theoretical value. The MSC-corrected data, however, correctly estimates the true signal variance and gives an experimental SNR that is in good agreement with the theoretical value.

3.2. Spectrograph performance comparison

Following the validation of using MSC correction to accurately compute Raman SNR from biological samples, this procedure was used to compare the performance of four preconfigured Raman spectrographs that can resolve relative wavenumber shifts between 950-1900cm⁻¹ at 785 nm excitation. The tested spectrographs all share similar spectral resolution and f-number, as seen by the summarized specifications in Table 2. Raman datasets were acquired from the biological milk standard according to the procedure outlined in Section 2.6.

Table 2. Specifications and Raman signal quality metrics of the tested spectrographs

Spectrograph	Specifications			Experimental		
	F-number	Slit Width (μm)	Grating Type	Resolution (cm ⁻¹)	SBR	SNR
System 1	f/2.2	200	Transmissive	8.9	0.19	61
System 2	f/2.2	100	Transmissive	8.7	0.20	45
System 3	f/1.8	100	Transmissive	8.3	0.11	23
System 4	f/2	75	Reflective	8.1	0.13	22

The mean spectra of the milk datasets from each spectrograph are presented in Fig. 5(a), and the background-subtracted mean spectra are shown in Fig. 5(b). Each system configuration used identical laser power, FOP, detector, and signal acquisition settings to collect Raman spectra, yet the four instruments demonstrate vastly different signal qualities in terms of throughput and

the relative proportion of Raman signal to background counts, termed the signal-to-background ratio (SBR). The general spectral line shape of preprocessed Raman data is equivalent for all spectrographs, as seen in Fig. 5(b), confirming that the Raman component of the signal was consistently detected. However, the total Raman intensity and SBR differed, leading to unique SNRs for each instrument when calculated according to Eq. (1). The optical specifications and experimental SBR/SNR metrics of each spectrograph are listed in Table 2.

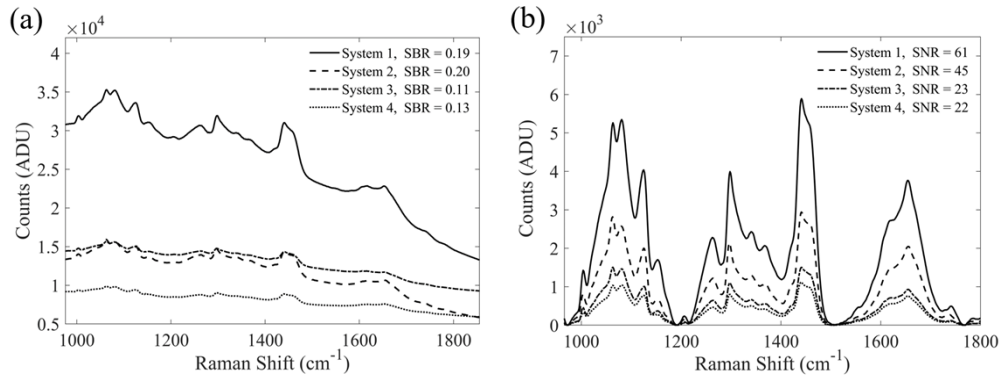


Fig. 5. (a) Mean milk spectra from the four Raman spectrographs with their associated signal-to-background ratios (SBR). (b) Background-subtracted mean spectra with their associated Raman signal-to-noise ratios (SNR) computed at 1440 cm^{-1} after MSC correction to remove photobleaching effects.

The noted discrepancy in Raman peak intensities between each system configuration indicates that the spectrographs held varying degrees of light throughput, resulting in a 6x gap in Raman intensity between the best- and worst-performing instruments. Since the acquisition procedure and biological standard ensures that an identical optical signal is delivered by the FOP, the differences in SBR must arise from differences in the spectrographs' specific optical designs. The optical throughput and SBR for each spectrograph led to unique Raman SNR values computed at 1440 cm^{-1} from the MSC-corrected milk datasets. For example, Fig. 5(a) shows that spectrograph 2 shared similar total signal counts to spectrograph 3 at 1440 cm^{-1} yet exhibited approximately 2x higher Raman signal and SBR. These compounding differences in total Raman signal counts and SBR led to a much higher SNR in Spectrograph 2 versus 3.

4. Discussion

Maximizing RS system sensitivity is essential to encourage the adoption of this technique for clinical diagnostics and other related fields that benefit from FOP-based Raman analysis of biologics. However, there are inherent challenges in comparing the signal quality of biological Raman spectra across system configurations. First, there is a strong dependence of probe orientation on the intensity of the detected spectrum, so the Raman signals used to compare system configurations are susceptible to the experimental variability induced by probe-sample positioning. Second, photobleaching effects are known to interfere with the experimental quantification of SNR from biological Raman spectra, resulting in non-biological samples being used to measure SNR and assess system performance [11,17]. Synthetic samples may not reflect the true response of the system to analyze organic materials, and they still do not solve the issue of FOP orientation dependence on the testing data.

By selecting a suitable biological sample that can produce a reliable and consistent Raman signal, this report establishes a methodology for unbiased performance assessment of RS system configurations for biomedical analysis. Results demonstrate that a dataset of 40 spectra can be

measured from this sample and corrected for photobleaching effects using MSC to obtain an accurate value for Raman SNR that was shown to agree with the theoretical SNR value. Future studies could explore the potential expedition of data collection through either increasing laser power or reducing the number of spectral measurements.

Dairy milk was chosen as the biological standard sample for multiple reasons. First, it is cheap and easily accessible. Second, it mimics many of the optical and chemical properties of tissues (i.e., diffuse reflectance, autofluorescence, organic biochemical composition) that manifest as a similar Raman spectrum to bulk tissue, as demonstrated in Fig. 4. Most significant to assessing the performance of RS systems that utilize FOPs, a homogenous liquid sample like milk eliminates any experimental variability related to probe-sample orientation. Although milk may not be the only suitable biological standard, it is certainly a practical option. It is important to remark that 2% reduced-fat milk may not be available in all countries. A dilution of raw dairy milk with water may be a strategy to tailor the spectral properties to better mimic that of tissues. However, this approach would dilute all organic components of milk and would not provide a way to control the SBR.

The MSC algorithm offers an efficient model-based correction step to eliminate photobleaching artifacts in the dataset and allows for the accurate calculation of SNR from biological Raman data, a known challenge in the field of biomedical RS. The Raman SNR was calculated from experimental data of the milk standard and a tissue specimen using Eq. (1) before and after applying MSC. Results showed that the MSC-correction approach, unlike the silent region method, correctly estimates the signal variance at the 1440 cm^{-1} Raman peak for both biological samples by providing an experimental SNR value that is in good agreement with the theoretically determined value. In addition, the suggested method for correcting photobleaching could have further implications for Raman data analysis. We believe that MSC could be a valuable tool in alternative situations where autofluorescence decay poses an issue, like when analyzing time-resolved RS data. Future studies could explore the use of other photobleaching correction methods and compare them with the MSC algorithm to determine the most effective method for correcting autofluorescence decay in Raman spectra. For example, exponential fitting of the fluorescent decay has been explored to correct photobleaching artifacts in microscopy images [32] but may be less reliable for RS data if the wavelength-dependent background decay rate is not uniform across the spectrum.

A spectrograph designed to measure biological Raman spectra requires high throughput and an optical design to minimize stray light. However, neither of these performance characteristics are obvious from the manufacturer-provided specifications alone. Since the method proposed here ensures that an identical optical signal is input to each spectrograph, it provides a direct assessment of their stray light performance through the SBR metric. Background signal counts increase for a spectrograph that is not optimally designed to minimize ambient light or laser scatter from reaching the detector. A Raman spectrum with lower SBR, as seen with Systems 3 and 4 in Fig. 5(a), contains a higher proportion of background counts relative to the usable Raman counts. In effect, a lower SBR increases total shot noise and, therefore, deteriorates Raman SNR. Similarly, differences in total optical throughput can be compared by the intensity of detected Raman peaks. Throughput is important to optimize for the inherently weak signals generated by Raman scattering and is dependent on the specific optical components used within the instrument; including the diffraction grating, entrance slit width, filters, and lenses. The background-subtracted signals in Fig. 5(b) demonstrate that each instrument also varied in the amount of detectable Raman light. The SBR/SNR signal quality metrics for each spectrograph are summarized in Table 2 and resulted in Systems 1 and 2 having a 3- and 2-fold increase in Raman SNR, respectively, compared to Systems 3 and 4. These results point to the importance of a proper methodology to compare RS system performance on biological samples, as the

manufacturer-provided specifications for these spectrographs would assume similar detection of the Raman spectra.

5. Conclusions

Assessing the performance of an FOP-based RS system for biomedical analysis is challenging due to signal instability in biological samples caused by photobleaching and experimental variances related to probe-sample orientation. The methodology presented here recommends dairy milk as a useful standard material to compare Raman systems as it provides a repeatable biological Raman spectrum by eliminating the dependence of probe orientation on the data. Photobleaching artifacts can then be corrected with MSC to accurately calculate Raman SNR. The utility of this approach was demonstrated by comparing four preconfigured spectrographs with similar optical specifications, yet results showed significant variations in SBRs and total Raman intensity. Ultimately, the unique system responses from each spectrograph led to vastly different Raman SNRs. While this methodology was used to compare Raman spectrographs in this report, it may be applied for alternative purposes. For instance, a researcher or engineer may use it to optimize the design of a custom-built spectrometer or to assess inter-probe variability when multiple FOPs are manufactured and expected to perform equivalently. It may also be used to determine how experimental parameters affect Raman SNR. Optimizing RS system performance for measuring biological samples with an FOP is crucial for a wide range of applications, including pharmaceuticals [33], food science [34], clinical diagnostics [3], and other biotechnologies [35]. Therefore, the methodology presented here has significant implications for researchers and industries that rely on RS for their analyses of biologics by presenting a consistent and unbiased method for assessing Raman system performance.

Funding. National Institute of Biomedical Imaging and Bioengineering (5R01EB028615-02).

Acknowledgements. The authors are grateful to Alec Walter, Jacob Hardenburger, and Richard Liao (Vanderbilt University, Nashville TN, USA) for the useful discussions and textual edits of this work.

Disclosures. There are no conflicts of interest that need to be declared.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. M. Schmitt, T. Mayerhöfer, J. Popp, I. Kleppe, and K. Weisshart, *Light-Matter Interaction* (Wiley-VCH, 2012), Chap. 13.
2. J. Popp, C. Krafft, and T. Mayerhöfer, "Modern Raman spectroscopy for biomedical applications: a variety of Raman spectroscopical techniques on the threshold of biomedical applications," *Optik & Photonik* **6**(4), 24–28 (2011).
3. E. Cordero, I. Latka, C. Matthäus, I. W. Schie, and J. Popp, "In-vivo Raman spectroscopy: from basics to applications," *J. Biomed. Opt.* **23**(07), 1 (2018).
4. H. P. S. Heng, C. Shu, W. Zheng, K. Lin, and Z. Huang, "Advances in real-time fiber-optic Raman spectroscopy for early cancer diagnosis: Pushing the frontier into clinical endoscopic applications," *Transl. Biophotonics* **3**(1), e202000018 (2021).
5. N. Sharma, N. Takeshita, and K. Y. Ho, "Raman spectroscopy for the endoscopic diagnosis of esophageal, gastric, and colonic diseases," *Clin. Endosc.* **49**(5), 404–407 (2016).
6. J. C. Day and N. Stone, "A subcutaneous Raman needle probe," *Appl. Spectrosc.* **67**(3), 349–354 (2013).
7. I. Pence and A. Mahadevan-Jansen, "Clinical instrumentation and applications of Raman spectroscopy," *Chem. Soc. Rev.* **45**(7), 1958–1979 (2016).
8. J. Zhao, H. Lui, D. I. McLean, and H. Zeng, "Integrated real-time Raman system for clinical in vivo skin analysis," *Skin Res. Technol.* **14**(4), 484–492 (2008).
9. M. Gong, H. Kim, J. Larsson, T. Methling, M. Aldén, E. Kristensson, C. Brackmann, T. Eschrich, M. Jäger, W. Kiefer, and A. Ehn, "Fiber-based stray light suppression in spectroscopy using periodic shadowing," *Opt. Express* **29**(5), 7232–7246 (2021).
10. N. Q. Dao, "Dispersive Raman spectroscopy, current instrumental designs," in *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, R.A. Meyers, ed. (Wiley, 2011).
11. I. J. Jahn, A. Grjasnow, H. John, K. Weber, J. Popp, and W. Hauswald, "Noise sources and requirements for confocal Raman spectrometers in biosensor applications," *Sensors* **21**(15), 5067 (2021).

12. J. Desroches, M. Jermyn, K. Mok, C. Lemieux-Leduc, J. Mercier, K. St-Arnaud, K. Urmeý, M. C. Guiot, E. Marple, K. Petrecca, and F. Leblond, "Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification," *Biomed. Opt. Express* **6**(7), 2380–2397 (2015).
13. M. D. Keller, E. Vargis, N. de Matos Granja, R. H. Wilson, M. A. Mycek, M. C. Kelley, and A. Mahadevan-Jansen, "Development of a spatially offset Raman spectroscopy probe for breast tumor surgical margin evaluation," *J. Biomed. Opt.* **16**(7), 077006 (2011).
14. H. A. Mustafa and O. Akkus, "Comparison of diffuse versus inverse spatially-offset Raman spectroscopy modalities for analyte detection through barriers," *Vib. Spectrosc.* **113**, 103228 (2021).
15. J. C. Lázaro, M. T. T. Pacheco, K. C. Rodrigues, C. J. de Lima, L. M. Moreira, A. B. Villaverde Jr, and L. Silveira, "Optimizing the Raman signal for characterizing organic samples: the effect of slit aperture and exposure time," *Spectroscopy* **23**(2), 71–80 (2009).
16. J. M. Smulko, N. C. Dingari, J. S. Soares, and I. Barman, "Anatomy of noise in quantitative biological Raman spectroscopy," *Bioanalysis* **6**(3), 411–421 (2014).
17. S. J. Barton and B. M. Hennelly, "Signal to noise ratio of Raman spectra of biological samples," *Proc. SPIE* **106854**, 698–708 (2018).
18. A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, "Introduction to quantum noise, measurement, and amplification," *Rev. Mod. Phys.* **82**(2), 1155–1208 (2010).
19. J. Rabinowitz, M. Rezaei, M. S. Park, C. L. Tan, M. Ulmer, and H. Mohseni, "When shot-noise-limited photodetectors disobey Poisson statistics," *Opt. Lett.* **45**(11), 3009–3012 (2020).
20. M. Hirsch, R. J. Wareham, M. L. Martin-Fernandez, M. P. Hobson, and D. J. Rolfe, "A stochastic model for electron multiplication charge-coupled devices—from theory to practice," *PLoS ONE* **8**(1), e53671 (2013).
21. C. Fajardo and E. Solarte, "Optical properties of a simple model of soft biological tissue," in *Proceedings of Journal of Physics: Conference Series*, (IOP Publishing, 2020), paper 012026.
22. W. Bachir, "Characterization of pasteurized milk in the near infrared range for construction of tissue-mimicking optical phantoms," *Opt. Mater.* **14**, 100154 (2022).
23. Z. M. Zhang, S. Chen, and Y. Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst* **135**(5), 1138–1146 (2010).
24. A. M. Macdonald and P. Wyeth, "On the use of photobleaching to reduce fluorescence background in Raman spectroscopy to improve the reliability of pigment identification on painted textiles," *J. Raman Spectrosc.* **37**(8), 830–835 (2006).
25. N. K. Afseth and A. Kohler, "Extended multiplicative signal correction in vibrational spectroscopy, a tutorial," *Chemometr. Intell. Lab.* **117**, 92–99 (2012).
26. S. Guo, A. Kohler, B. Zimmermann, R. Heinke, S. Stöckel, P. Rösch, J. Popp, and T. Bocklitz, "Extended multiplicative signal correction based model transfer for Raman spectroscopy in biological applications," *Anal. Chem.* **90**(16), 9787–9795 (2018).
27. K. H. Liland, A. Kohler, and N.K. Afseth, "Model-based pre-processing in Raman spectroscopy of biological samples," *J. Raman Spectrosc.* **47**(6), 643–650 (2016).
28. A. S. Luna, A. P. da Silva, C. S. da Silva, I. C. Lima, and J. S. de Gois, "Chemometric methods for classification of clonal varieties of green coffee using Raman spectroscopy and direct sample analysis," *J. Food Compos. Anal.* **76**, 44–50 (2019).
29. H. Vašková and M. Bučková, "Spectroscopic measurement and analysis of fat in milk," in *Proceedings of 26th DAAAM International Symposium on Intelligent Manufacturing and Automation*, B. Katalinic, ed. (DAAAM International, 2015), pp. 365–370.
30. J. Reiner, K. Protte, and J. Hinrichs, "Investigation of the applicability of Raman spectroscopy as online process control during consumer milk production," *ChemEngineering* **4**(3), 45 (2020).
31. H. Martens, A. Kohler, N. K. Afseth, J. P. Wold, M. Harsleth, I. Berget, T. Ådnøy, M. Skaugen, T. Isaksson, G. Vegarud, and A. Criscione, "High-throughput measurements for functional genomics of milk," *J. Anim. Feed Sci.* **16**(Suppl. 1), pp.172–189 (2007).
32. N. B. Vicente, J. E. D. Zamboni, J. F. Adur, E. V. Paravani, and V. H. Casco, "Photobleaching correction in fluorescence microscopy images," in *Proceedings of Journal of Physics: Conference Series* (IOP Publishing, 2007), paper 012068.
33. K. A. Esmonde-White, M. Cuellar, C. Uerpmann, B. Lenain, and I. R. Lewis, "Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing," *Anal. Bioanal. Chem.* **409**(3), 637–649 (2017).
34. K. Chao, S. Dhakal, J. Qin, M. Kim, and Y. Peng, "A 1064 nm dispersive Raman spectral imaging system for food safety and quality evaluation," *Appl. Sci.* **8**(3), 431 (2018).
35. P. C. Ashok, M. E. Giardini, K. Dholakia, and W. Sibbett, "A Raman spectroscopy bio-sensor for tissue discrimination in surgical robotics," *J. Biophotonics* **7**(1-2), 103–109 (2014).